2280 J. Acoust. Soc. Am. 91 (4), Pt 1, April 1992

An acoustical study of Korean monophthongs produced by male and female speakers

Byunggon Yang
Department of English, College of Hmanities, Dongeui University, 24 Kaya-dong, Pusanjin-gu, Pusan 614-714, Korea

(Received 21 August 1990; accepted for publication 29 November 1991)

The first three formants of the ten Korean monophthongs were studied. The vowels were produced by 20 male and female speakers in an |hVd ɑ | (V = vowel) context. Then the male-female variations in the Korean formant data were examined. The uniform scaling method resulted in less than a 6% difference between the scaled and reference data. A regression analysis showed that the slope of a regression equation with a small intercept near the origin could be used for the uniform scale factor.

INTRODUCTION

Previous studies have shown that formant frequencies are important factors in determining the perceptual quality of a vowel. Hence, each vowel phoneme has its characteristic formant patterns within a given language. However, spectrographic studies of vowels produced by different speakers of a given language show great variation in formant frequencies dependent on age and gender. This speaker variation causes many problems in determining the acoustic correlates of the vowel qualities of a language. These sources of speaker variation are divided into linguistic and nonlinguistic factors (Traunmüller, 1988). Linguistic factors are dialectal, sociolectal, idiolectal, and phonostylistic differences. Nonlinguistic factors include the differences in age, gender, and the emotional state of the speakers. Since the vowel formants depend on the length and shape of the vocal tract, a major source of nonlinguistic speaker differences is the individual speaker's vocal tract anatomy (Fant, 1960, 1968; Nordström and Lindblom, 1975; Nordström, 1975).

The goal of factoring out these nonlinguistic factors is to establish a "pure," linguistically relevant acoustic specification of the vowel qualities of any given language. This procedure has been called "normalization " (Fant, 1968). Various proposals for the normalization of vowel qualities have been explored, but the problem has not yet received its definitive solution. Some proposals are articulatorily based (Nordström and Lindblom, 1975; Fant, 1975), while others are auditorily based (Miller et al., 1980; Miller, 1989; Syrdal and Gopal, 1989), involving the transformation of the formant values into auditory units. Auditory normalizations are based on the belief that there is a constancy in the spatial pattern of excitation in the peripheral auditory system for a given vowel produced by different speakers. In this study the discussion is limited to the two articulatorily based proposals: uniform and nonuniform scaling methods. Nordström and Lindblom (1975) proposed a uniform scaling based on an estimate of the length of the speaker's vocal tract. According to their claim, all observed speaker-dependent formant frequency differences can be satisfactorily accounted for soley in terms of vocal tract length. Assuming that overall vocal tract length variations affect all formant frequency values by the same scale factor (that is, in a uniform way), the same scale factor was applied to all formants. Therefore, the term "uniform scaling" was used. It involved estimating the total length of a subject's vocal tract from an average of F3 in vowels with F1 greater than 600 Hz. Because the length of the speaker's vocal tract is inversely related to formant frequency, the ratio of the length of the average male vocal tract (Lm) to the average female vocal tract length (Lf) can be written:

$$k = Lf/Lm = F3mav/F3fav \, , \qquad (1)$$

in which F3mav and F3fav indicate an average of the third male and female formant values, respectively. Then, the normalized nth female formant frequency is denoted as Fnf (scaled) and can be determined according to

$$Fnf(scaled) = kFnf, \qquad (2)$$

On the other hand, Fant (1968) observed that female speakers have proportionately shorter pharynges and argued that the scaling of formant values must be nonuniform. In other words, it must consider not only differences in the overall vocal tract length between male and female speakers but also the complex formant-cavity relationships. Therefore, Fant (1975) recommended using scale factors that are both vowel and formant specific. His method applies a different scale factor to each individual vowel and individual formant category. A tentative reference table of "universal" scale factors was proposed based on the statistical study of six languages: Swedish, American English, Danish, Estonian, Dutch, and Serbo-Croatian (Fant, 1975, p.5, Table I-A-I). Because the table was derived from the average differences of male and female formant frequencies only for those six languages, it may not be proper to use those factors in

----

normalizing the Korean data.

In this paper, the first three formants of the ten Korean monophthongs produced by 20 male and female speakers were studied while controlling the linguistic factors as homogeneously as possible in each group. Second, the male-female variations in the Korean data were examined. Third, the data were scaled using the uniform scaling method. Fourth, regressional analyses were conducted on formant values of male and female groups.

I. METHOD

A. Subjects and speech samples

A total of 20 subjects were chosen from a larger group participating in recording and listening sessions at the University of Texas at Austin. They formed two groups: ten Korean males and ten Korean females. Subjects were students attending the University of Texas at Austin and all had normal hearing and health. All the Korean subjects spoke Standard Korean. Two screening instruments were used to make each group linguistically homogeneous.

First, subjects were grouped homogeneously on the basis of collected information from a questionnaire. The questionnaire included subjects' name, age, sex, height, native language, fluency in other languages, dialect, and history of speech and hearing disorders. Second, peer judgment was employed to screen out those subjects who had different dialects. Five peers in each male and female group were randomly chosen from among the subjects. Then the ten peers were asked to listen to four sets of tokens consisting of the vowels [i e ɑ u]. Each set is composed of different male and female subjects saying the same token. In the listening session, the peers put a check mark on each token that sounded different from their own dialect. All the marks were counted to find four peers (two -males and two females) who had the fewest marks. Finally, marks by the four peers were used to reject those subjects who had more than 80% of the total tokens perceived as a different dialect.

The speech samples consisted of 52 Korean words. Each Korean vowel occurred in an |h(V)dɑ| (V = vowel) context. In this context, the following vowel formant can be easily identified because the |h| noise on the spectrogram shows similar patterns of the following vowel formants. Ten Standard Korean monophthongs studied were / ɑ e ɛ o ø u ü i ʌ ɨ/. These ten Korean vowels appeared five times in random order. Later, three out of the five productions of each vowel were randomly chosen for the average data set, avoiding unnatural tokens at the beginning and ending of the recording.

B. Procedures

The recording was done in a sound-proof booth in the Phonetics Lab of the University of Texas at Austin (UT). The experimenter asked the subjects to produce each word at a normal rate and as naturally as possible. Speech samples were picked up by a condenser microphone at a distance of about 15 cm. A professional cassette recorder (Model CD-330) was used for recordings. The recording level was carefully monitored by the experimenter throughout the recording session to avoid ultraweak or overloaded signals. The recording took 2-3 min per subject.

The recorded samples were analyzed using the VAX computer in the UT Phonetics Lab. The KLSPEC software package written by D. H. Klatt and modified by J. Lame was used to interactively examine, measure, and analyze the recorded samples. The samples were played from a Yamaha (model KX-800U) cassette deck. The input samples were low pass filtered at 4 kHz and digitized at a 10-kHz sampling rate. Spectrograms were made using a 256-point discrete

Fourier transform (DFT) analysis with a 6.4-ms Hamming window once every millisecond. The dynamics of the vowel formant pattern made it difficult to find a consistent time point for spectrum analysis. Therefore, the author calculated the total duration from the vowel onset and offset time points. Each time point was determined by adding one-third of the total duration to the vowel onset point.

C. Formant decisions

In this study, a spectrogram of each vowel was used for the first reference and final decision on formant values. First, formant values on the spectrogram were estimated by drawing a pencil line through the center of each formant band with a ruler. Then, visual estimates were compared with two DFT harmonic spectra (average envelope and LPC envelope) estimates. The output of the software package listed estimates of three or more feasible formant values along with amplitude levels. These methods almost always converged. However, when formant values of the same vowel for a subject showed a difference of more than 500 Hz, they were rejected under the assumption that the subject produced a different target vowel. For reliability, the measurements were checked by an independent observer. There was an agreement of more than 97% in the formant decision.

II. RESULTS AND DISCUSSION

The average values of the first three formants (FI, F2, and F3) and their standard deviation (s.d.) of the Korean monophthongs are listed in Tables I and II. As is shown in the tables, the deviation within males is generally lower than that of females. Figure 1 illustrates the vowel space of males and females in which adjacent vowel points are connected peripherally. Phonemes are given near various symbols. The

TABLE I. Average values of the first three formants (F1,F2,F3) and their standard deviation (s.d.F1, s.d.F2, s.d.F3) of the Korean female speakers.

Vowel	F1	s.d.F1	F2	s.d.F2	F3	s.d.F3
ɑ	986	107	1794	108	2957	227
ɛ	677	108	2285	169	3063	141
e	650	113	2377	77	3068	117
i	344	48	2814	168	3471	177
o	499	60	1029	143	3068	159

ø	602	109	2195	152	3013	132
u	422	83	1021	139	3024	138
ü	373	62	2704	95	3222	108
ʌ	765	125	1371	108	3009	183
ɨ	447	68	1703	106	2991	173

TABLE II Average values of the first three formants (F1,F2,F3)and their standard deviation (s.d.F1, s.d.F2, s.d.F3) of the Korean male speakers.

Vowel	F1	s.d.F1	F2	s.d.F2	F3	s.d.F3
ɑ	738	81	1372	124	2573	127
ɛ	591	75	1849	106	2597	110
e	490	105	1968	150	2644	94
i	341	29	2219	176	3047	146
o	453	47	945	134	2674	156
ø	459	69	1817	163	2468	134
u	369	43	981	141	2565	173
ü	338	30	2114	140	2729	213
ʌ	608	76	1121	110	2683	145
ɨ	405	37	1488	176	2497	80

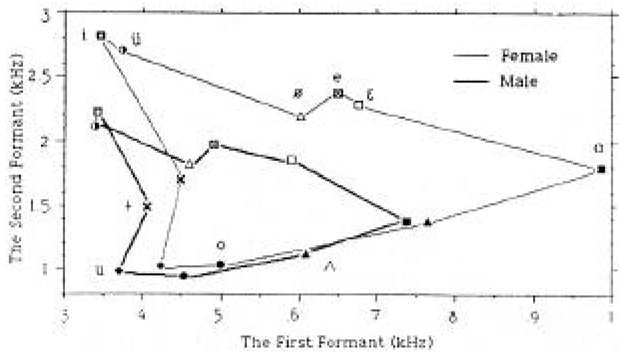
same symbol indicates the same phoneme. The two vowel spaces appear triangular. A thicker line connects male vowels, while a thinner line connects female vowels. The vowel spaces show some systematic relationship. It implies a trend toward higher formant frequencies for female speakers. This tendency may be largely due to nonlinguistic factors because linguistic factors in this experiment were homogeneously controlled in each group as much as possible. To examine that relationship, the percent difference (Diff.) in the male and female formant frequencies across vowel formant was calculated by

$$\text{Diff.}(\%) = \{ (Fnf - Fnm) / Fnm \} \times 100, \qquad (3)$$

in Which Fnf denotes the *n*th female formant value. The average Diff. across all the vowels comes out to 18% which equals the average difference of F1. In F2 and F3 the Diffs. are 20% and 17%, respectively. In addition the greatest variation within each formant was in F1 (s.d.= 12%) followed by F2 (s.d.= 9%) and F3 (s.d.= 3%). This implies that male data may be uniformly shifted to the female data or vice versa. Thus the uniform scaling method was applied to the Korean data. The uniform scale factor is 0.87 estimated from Eq. (1) .This ratio corroborates the study by Chiba and Kajiyama (1941) who estimated overall vocal tract length assigning the relative numbers of 1.0 to males and 0.87 to females. The data were scaled by Eq. (2). Then, a numerical criterion was used to see how closely the female data were

FIG. 1. Male and fe male vowel space of ten Korean monophthongs. x axis shows the first formant frequency in kiloHertz (kHz) while y axis indicates

the second formant frequency.



scaled to the male reference data (Fant, 1975). For that purpose, Fnf in Eq. (3) was replaced by Fnf (scaled). The uniform scaling method had an average Diff. of less than 6% across all the vowels. Specifically, there was more than a 10% Diff. in F1 of the uniformly-scaled data, but less than a 5% Diff. was observed in back vowels. In F2 the point vowels [i ɑ u] showed a 10% Diff. while the others were less than 10%. In the vowel /i/ a perfect match occurred. In F3t an average Diff. of 2% was observed which stands to reason since the scale factor was derived from the ratio of male to female averages of F3.

The question arises as to whether the nonuniform scaling greatly improves the uniform scaling method in the Korean data. Fant (1975) suggested that the gender variation can be best resolved by formant-number and vowel-specific scale factors. Certain Korean vowels show greater male-female difference. For example, F1 in the vowel [0] shows a

34% Diff. while that of the vowel [i]shows less than a 1% Diff. Since the scale factors in the nonuniform methods were derived from the average variation of the six European languages they may not be, proper to use in the scaling of the Korean data. However t one can reason that if one employs a formant-number-specific method, the average Diff. would improve around 1% because there was about 1% difference among F I' F 2t and F3 in the previous analysis. Moreover if both the formant-number and vowel-specific scale factors are used, then it will improve less than 6%.
Finally regressional analyses were conducted to find regression equations or the best-fitting lines for the relationship between all the female formant values (Fnf) and those of males (Fnm) or vice versa:

Fnm =O.85 Fnf- 5/ Fnf =1.18Fnm + 19 (r^2 /0.99)

The slope 0.85 is similar to the average of F3f/ F3m (r^2= 0.87) for the uniform scaling method. With small intercepts, both equations imply that one may expect a good fit with a regression equation through the origin (zero intercept). Thus the uniform scale factor can be easily determined through the regression. The r^2 indicates that female formant values can be accurately predicted from male values or vice versa. In addition, a regression analysis between female formant frequency and the male-female difference in percent revealed a strong correlation (Fnf against Diff., r < 0.78). This corroborates Fant's observation (1975) that the gender difference in F1 and F2 increases with formant frequency. O'Leary (1989) has also observed a strong correlation between scale factors obtained with Fant's scaling method and female formant values. Fant's scale factors were higher than the average for high vowels while those for lower vowels were lower . Using the regression equations relating Fant’ s scale factors to female formant frequencies, O'Leary scaled female formant frequencies to match more closely those of males.
In conclusion the first three formants of the ten Korean monophthongs were studied. Because the linguistic factors were controlled to be as homogeneously as possible in each group, the male-female difference was attributed to nonlin-, guistic factors. The uniform scaling method resulted in less than a 6% difference between the scaled and reference data. A regression analysis showed that the slope of a regression

equation with a small intercept near the origin could be used
for the uniform scale factor .

Chiba, T., and Kajiyama, M. (1941). The vowel-its Nature and Structure (Kaiseikan, Tokyo) .

Fant, G. (1960). Acoustic Theory of Speech Production (Mouton, 's-Gra venhage, The Netherlands).

Fant, G. (1968). Analysis and Synthesis of Speech Processes," in Manual Of Phonetics, edited by B. Malmberg (North Holland, Amsterdam), pp. 243-253.

Fant, G. (1975). '.Speech Production," STL-QPSR 2-3,1-19.

Miller, J. D., Engebretson, A. M., and Vemula, N. R. (1980). "Vowel Normalization: Differences between vowels spoken by children, women, and men," J. Acoust. Soc. Am. Suppl. 1 68, 833.

Miller, J.D. (1989). "Auditory-perceptual interpretation of the vowel,"J. Acoust. Soc. Am. 85,2114-2134.

Syrdal, A. K., and Gopal, H. S. (1989). 'A perceptual model of vowel rec ognition based on the auditory representation of American English vow els," J. Acoust. Soc. Am. 79, 108~1100.

Nordström, P. E. (1975). ., Attempts to simulate female and infant vocal tracts from male area functions," STL-QPSR 2-3, 20-33.

Nordström, P-E., and Lindblom, B. (1915). "A normalization procedure for vowel formant data," Paper 212 at the International Congress of Phonetic Sciences in Leeds, August.

O'Leary, M. (1989). "Scaling as a linear function of formant values" (un published).

Traunmüller, H. (1988). "Paralinguistic variation and in variance in the characteristic frequencies of vowels," Phonetica 45, 1-29.